Molecular Phylogenetics and Evolution 48 (2008) 270–280

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Molecular evaluation of nominal species in the *Ceratitis fasciventris*, *C. anonae*, *C. rosa* complex (Diptera: Tephritidae)

M. Virgilio [a,b,*], T. Backeljau [b,c], N. Barr [d], M. De Meyer [a]

[a] Royal Museum for Central Africa, Leuvensesteenweg 13, 3080 Tervuren, Belgium
[b] Royal Belgian Institute of Natural Sciences, Vautierstraat 29, 1000 Brussels, Belgium
[c] Evolutionary Ecology Group, University of Antwerp, Groenenborgerlaan 171, B-2020 Antwerp, Belgium
[d] PDDML-APHIS-USDA, Moore Air Base, 22675 N. Moorefield Road, Edinburg, TX 78541, USA

ARTICLE INFO

Article history:
Received 26 November 2007
Revised 20 March 2008
Accepted 9 April 2008
Available online 23 May 2008

Keywords:
Fruit flies
Pest species
Species complex
Cryptic species
DNA taxonomy

ABSTRACT

Ceratitis fasciventris, *C. anonae* and *C. rosa* form a complex of economically important fruit fly pests infesting a variety of crops in African countries. Hitherto only adult males of these species can be distinguished easily by morphological characters. Other stages cannot, and for some taxa the taxonomic interpretation and species boundaries remain unclear. In order to clarify phylogenetic relationships and taxonomic status of these species, sequences of mitochondrial (16S, COI, ND6) and nuclear markers (*period*, ITS1) were analysed in specimens of the three morphospecies throughout the distribution of the complex. Maximum likelihood trees did not recover monophyletic groups corresponding to the morphospecies. Conversely, ND6 and COI divided West African *C. fasciventris* specimens in two consistent and bootstrap supported clades, involving specimens from Benin and from Mali/Ivory Coast, while the nuclear gene fragments *per* and ITS1 recovered a well-supported clade corresponding to *C. fasciventris* from Kenya/Uganda. Hence, the phylogenetic relationships and taxonomic interpretation of the complex appear more intricate than previously hypothesised. The current molecular data do not allow to identify *C. fasciventris*, *C. anonae* and *C. rosa* as distinct phylogenetic species but rather suggest that the morphospecies *C. fasciventris* is itself a complex of cryptic taxa.

1. Introduction

The Afrotropical fruit fly genus *Ceratitis* MacLeay (Diptera: Tephritidae) comprises 89 species grouped in six subgenera (De Meyer, 2005; Barr and McPheron, 2006). The genus includes economically relevant agricultural pests (Malacrida et al., 2007) and a number of species complexes (De Meyer, 2005). Among them, the *Ceratitis fasciventris*, *C. anonae* and *C. rosa* complex (hereafter referred to as FAR complex) consists of polyphagous pests of horticultural crops, belonging to diverse plant families including Annonaceae, Myrtaceae, Rosaceae and Rubiaceae (De Meyer et al., 2002; Copeland et al., 2006). The host spectra of the species within the FAR complex have some considerable overlap, even if each species also infests a number of unique hosts (De Meyer et al., 2002; Copeland and Wharton, 2006; Copeland et al., 2006). The three species of the complex belong to the subgenus *Pterandrus* and are widespread in a number of African countries. *C. fasciventris* and *C. anonae* occur sympatrically in both East and West Africa, while *C. rosa* is more restricted to southern and eastern Africa where its distribution partially overlaps with that of *C. fasciventris* but not with *C. anonae* (see Copeland et al., 2006; De Meyer, 2001a).

Due to the risks of accidental introductions of potentially highly invasive *Ceratitis* pests (Hancock, 1989; Duyck et al., 2004; Copeland et al., 2006; Malacrida et al, 2007; De Meyer et al., 2008) there is an increasing need to reliably identify the invasive fruit flies before they can establish themselves (Barr et al., 2006). Hitherto, the identification of species of the FAR complex is based on morphological characters of adults (De Meyer, 2001b, 2005; De Meyer and Freidberg, 2006). The distinction of males of the three species relies on differences in leg ornamentation patterns (De Meyer and Freidberg, 2006). These differences are discrete and consistent, although some variability is observed between western and eastern African populations of *C. fasciventris* (De Meyer, 2001b). Females of *C. anonae* can, however, only be distinguished from the other two species by minute differences in their leg pilosity colour, while the distinction between females of *C. rosa* and *C. fasciventris* relies on even more subtle morphological characters (De Meyer, 2001b). Immatures such as larvae or pupae, the stages routinely encountered by quarantine officers in interceptions, are currently impossible to distinguish from each other. Hence, molecular markers may provide additional, and possibly more straightforward, tools to identify these pests and to investigate their phylogenetic

* Corresponding author. Address: Royal Museum for Central Africa, Leuvensesteenweg 13, 3080 Tervuren, Belgium. Fax: +32 26274141.
E-mail address: m.virgilio@unibo.it (M. Virgilio).

1055-7903/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2008.04.018

relationships. Because of the importance of developing appropriate pest detection, diagnostic and management strategies for these species it is important to evaluate the current morphospecies definitions, by characterizing the genetic variation within, and differentiation between, the species, as well as by assessing their evolutionary relationships.

Several earlier studies already dealt with the phylogeny and population genetics of species of the FAR complex. Baliraine et al. (2004), for example, showed that microsatellites allow to distinguish *C. fasciventris* from *C. rosa* in part of their distributional area (including Southern and Eastern African locations for *C. rosa* and East African locations for *C. fasciventris*). Barr and McPheron (2006) investigated the phylogeny of 32 species of the genus *Ceratitis* by sequence analysis of mitochondrial and nuclear gene fragments. The results confirmed that the subgenus *Pterandrus* comprises the sections A and B previously described by Hancock and White (1997) through morphological analyses. Both morphological (Hancock and White, 1997) and genetic analyses (Barr and McPheron, 2006) then placed the FAR complex in *Pterandrus* section A. Preliminary analyses of intraspecific sequence variation of the *NADH-dehydrogenase* subunit 6 locus (ND6) failed, however, to recover *C. fasciventris*, *C. anonae* and *C. rosa* as monophyletic taxa (Barr and McPheron, 2006). Accordingly, a pilot study investigating the reliability of the COI barcode region (as defined in Hebert, 2003) in the molecular diagnosis of Tephritids could not consistently separate several specimens of the species of the FAR complex (Virgilio, unpublished data). These results suggest that taxon sampling may heavily affect the accuracy of inferences drawn about the FAR complex and represents an important factor to consider in the analysis of closely related Tephritid species. Hence, despite earlier molecular analyses, the phylogenetic relationships and the taxonomic status of taxa in the FAR complex still remain to be clarified.

A number of studies investigated the possibility of developing molecular tools for the diagnosis of species belonging to the FAR complex. Douglas and Haymer (2001) reported ITS1 amplicon size differences within a number of *C. rosa* samples and proposed this locus as a possible tool either for population assignments and/or for species identification. Further PCR–RFLP analyses (Barr et al., 2006) showed that amplicon size differences of ITS1 may be used to distinguish *C. rosa* from both *C. fasciventris* and *C. anonae*. However, these conclusions relied heavily on samples from Kenya and the authors remarked that they may only apply to individuals from this geographical area. Hence, there is an urgent need to evaluate if patterns of ITS1 differentiation are geographically consistent across the whole distributional range of the complex. Based on previous results, ITS1 may also be suitable to reconstruct phylogenetic relationships of the species of the FAR complex.

In order to determine if the morphospecies descriptions are congruent with a phylogenetic classification and/or distinguishable using molecular characters we examined DNA sequences of multiple gene fragments, using samples from across the species' native ranges in Africa. The objectives of this study were to (a) verify whether the three currently recognized morphospecies of the FAR complex are phylogenetically distinct groups, (b) compare the results of alternative molecular markers with those already applied in previous studies (e.g. ND6 and ITS1), this time considering samples from the whole distributional area of the species, and (c) verify if ITS1 is a reliable diagnostic marker to identify taxa of the FAR complex.

2. Materials and methods

2.1. Sampling design and laboratory analyses

We analysed FAR specimens from the collections of the Royal Museum for Central Africa (Tervuren, Belgium). Specimens were selected in order to cover as much as possible the distributional range of the three species, including West and East Africa for *C. anonae* and *C. fasciventris* and East and South Africa for *C. rosa* (Fig. 1 and Table 1). Whenever possible, we included locations where representatives of FAR have been reported to occur in sympatry (e.g. Kenyan Central and Western Highlands provinces, Copeland et al., 2006). DNA was extracted from 56 specimens of the three morphospecies. Sequences were produced for the mitochondrial loci 16S rDNA (16S, GenBank Accession Nos. EU276759–EU276814), COI (Accession Nos. EU276645–EU276700), ND6 (Accession Nos. EU276702–EU276757) and for the nuclear locus *period* (*per*, Accession Nos. EU276816–EU276871).

Sequence variation among ITS1 copies was investigated considering variability among and within morphospecies and within specimens. Inter- and intra-specific variation was analysed in 50 FAR specimens. A subset of these specimens ($n = 21$) was used to quantify intra-individual ITS1 variability. Two to nine cloned ITS1 sequences were obtained for each specimen of the subset (Table 2). A total of 96 ITS1 sequences were produced (Accession Nos. EU276873–EU276967).

Sequences of the 16S, COI, ND6, per and ITS1 gene fragments were also obtained from a specimen of *C. rubivora* (subgenus *Pterandrus* section A, Barr and McPheron, 2006, Accession Nos. EU276758, EU276644, EU276701, EU276815, EU276872), which was used as outgroup (see below).

DNA was extracted from both pinned and ethanol preserved specimens using the DNeasy Blood and Tissue Kit (Qiagen) and following the manufacturer's protocol. All sequence fragments were amplified using primers and protocols described in Barr and McPheron (2006) and Barr et al. (2006). PCR products were purified by means of GFX purification columns (GE Healthcare), subjected to sequencing reactions using the BigDye cycle sequencing kit (Applied Biosystems) and sequenced in both directions with an ABI Prism 3100 Genetic Analyzer (Applied Biosystem). Before sequencing, ITS1 PCR products were cloned into TA vector pCR4 (Invitrogen).

2.2. Phylogenetic analyses

Nucleotide sequences were aligned using the ClustalW algorithm included in the Bioedit 7.0 software package (Hall, 1999). Heterozygotes for the *per* fragment were inferred directly from chromatograms and reported following the IUPAC conventions. Polymorphism analyses were performed using Mega 3.1 (Kumar et al., 2004) and DnaSP 4.10 (Rozas et al., 2003). Assuming mutation–drift balance for species, departures from neutrality for individual loci were assessed using Tajima's D statistic (Tajima, 1989) and tested for significance through the permutational procedures implemented by Arlequin 2.000 (Schneider et al., 2000). Probability values of repeated Tajima's D tests were corrected for Type I errors using the False Discovery Rate procedure (Benjamini and Hochberg, 1995). The amount of phylogenetic signal provided by the different markers was quantified by means of likelihood mapping (Strimmer and von Haeseler, 1997) as implemented by TREE-PUZZLE 5.2 (Schmidt et al., 2002). For this analysis, all the possible quartets were considered and transition/transversion ratios, nucleotide frequencies and substitution rates estimated from the datasets. The PHYML software package (Guindon and Gascuel, 2003) was used for maximum likelihood (ML) tree reconstruction. Substitution rate categories were set to four and gamma shape parameters, proportions of invariable sites, transition/transversion ratios and nucleotide frequencies were estimated from the datasets. Branch support in the ML trees was tested by means of 100 bootstrap replicates. Both the quartet tests and ML analyses were implemented by using the evolutionary models selected by the Akaike information criterion of MrModeltest (Nylander, 2004). The HKY model (Hasegawa et al., 1985) was used for 16S, ITS1,

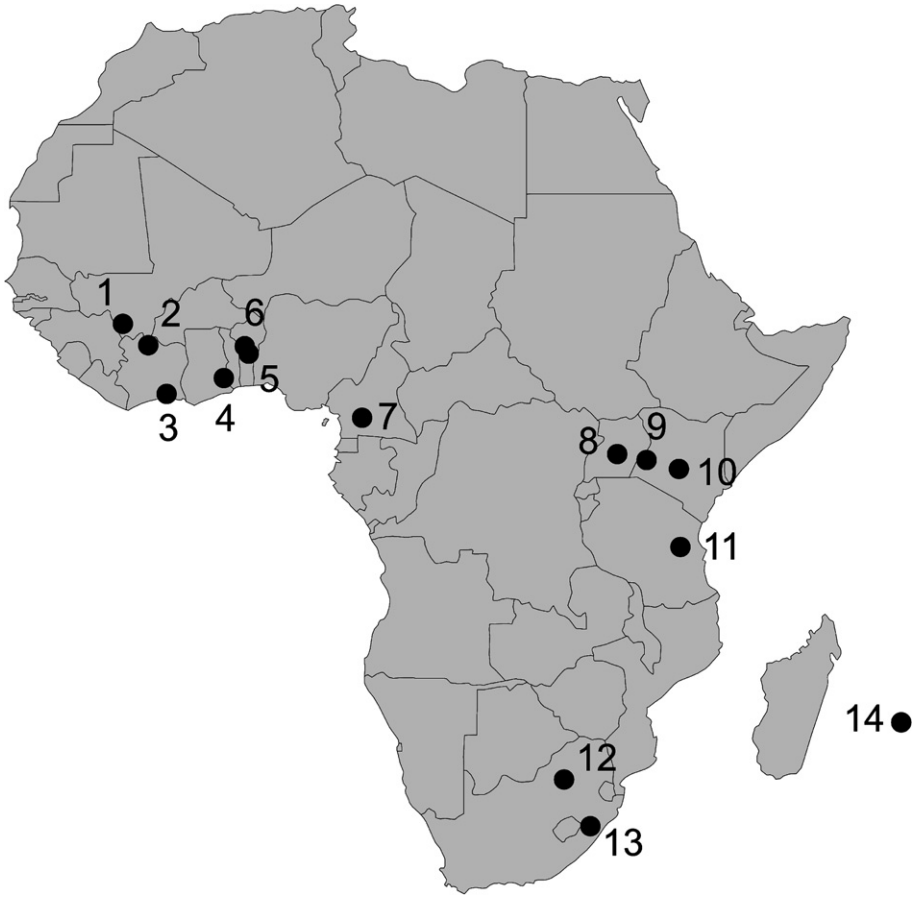

Fig. 1. Map of the African continent showing the geographical coverage of specimens analysed. Sites are numbered according to Table 1.

Table 1
Geographic distribution (see also Fig. 1) and number of specimens of *C. fasciventris*, *C. anonae* and *C. rosa*

Site	Country	Location	*C. fasciventris*		*C. anonae*		*C. rosa*	
			COI, ND6, 16S, *per*	ITS1	COI, ND6, 16S, *per*	ITS1	COI, ND6, 16S, *per*	ITS1
West								
1	Mali	Madina	2	1				
2	Ivory Coast	Korhogo				3		
3	Ivory Coast	Azaguié	3	2				
4	Ghana	Ashanti				3		
5	Benin	Koro	4	2	4	4		
6	Benin	Kakara		2				
7	Cameroon	Bikoh			5	3		
East								
8	Uganda	Kisule	2	2	4	3		
9	Kenya	Western Highlands	9			6	4	
10	Kenya	Central Highlands	2	4	3			
11	Tanzania	Morogoro		1			5	4
South								
12	South Africa	Pretoria						6
13	South Africa	Ferncliffe, Natal					4	
14	Reunion	CIRAD Lab					5	4
		Total	22	14	16	22	18	14

Specimens were used to produce multiple COI, ND6, 16S, *per* and/or ITS1 sequences.

ND6 and *per*, whereas the GTR model (Tavaré, 1986) was used for COI. Under conditions thought to be typical of most phylogenetic analyses, bootstrap support (BS) values in majority rule consensus trees provide biased but highly conservative estimates of the probability of correctly inferring the corresponding clades (Hillis and Bull, 1993). In order to reduce the risks of misinterpreting weakly supported clades, we considered only groups with BS $\geqslant$ 70% as well-supported clades (Hillis and Bull, 1993). Shimodaira and Hasegawa (1999) tests were used to verify the congruence between (1) ML tree topologies and the current morphological taxonomic interpretation of the FAR complex and (2) ML tree topologies provided by the different markers. The ITS1 gene fragment was not included in this second set of tests as the ML tree obtained for this fragment was obtained using a different subset of FAR specimens. The SH test routine of PAUP (Swofford, 2002) was used to compare (1) likelihood scores (LS) of unconstrained trees from the COI, ND6, 16S, ITS1 and Period gene fragments and LS of trees constrained for the monophyly of all the specimens of either *C. anonae*, *C. rosa* or *C. fasciventris* and (2) LS of unconstrained trees and trees constrained for the topologies of COI, ND6, 16S and *per*. In order to make the condition of the SH test less stringent, by reducing the weight of weakly supported clades, we considered only nodes with bootstrap support higher than 70% (see Hillis and Bull, 1993). Therefore, 70% majority rule consensus trees were built using PAUP (Swofford, 2002) and used to compare topologies. Probability values of repeated SH tests were corrected for Type I errors using the False Discovery Rate procedure (Benjamini and Hochberg, 1995). ITS1 sequences were compared with those previously published by Douglas and Haymer (2001) and Barr et al. (2006). These sequences (Accession Nos. AF189689, AF189690, DQ645953, DQ645954) were considered to represent the three different ITS1 length variants described so far in *C. fasci-*

Table 2
Number of specimens of *C. fasciventris*, *C. anonae* and *C. rosa* considered in the analysis of intra-individual variability of ITS1

Species	Specimen	Country	Location	No. of cloned seq.	GenBank Accession Nos.
C. fasciventris	1	Uganda	Kisule	2	EU276877–EU276878
	2	Benin	Koro	3	EU276881–EU276883
	3	Ivory Coast	Azaguié	2	EU276887–EU276888
	4	Ivory Coast	Azaguié	2	EU276889–EU276890
	5	Mali	Madina	4	EU276891–EU276894
C. anonae	1	Ghana	Ashanti	3	EU276898–EU276900
	2	Ghana	Ashanti	2	EU276901–EU276902
	3	Uganda	Kisule	2	EU276913–EU276914
	4	Benin	Koro	2	EU276917–EU276918
	5	Benin	Koro	2	EU276920–EU276921
	6	Benin	Koro	3	EU276922–EU276924
C. rosa	1	Tanzania	Morogoro	3	EU276925–EU276927
	2	Tanzania	Morogoro	3	EU276928–EU276930
	3	Tanzania	Morogoro	3	EU276931–EU276933
	4	Tanzania	Morogoro	2	EU276934–EU276935
	5	Reunion	CIRAD Lab	2	EU276939–EU276940
	6	South Africa	Pretoria	5	EU276941–EU276945
	7	South Africa	Pretoria	5	EU276947–EU276951
	8	South Africa	Pretoria	4	EU276952–EU276955
	9	South Africa	Pretoria	4	EU276956–EU276959
	10	South Africa	Pretoria	8	EU276960–EU276967

For each specimen, sampling location, number of cloned ITS1 sequences and GenBank accession numbers are indicated.

ventris, *C. anonae* and *C. rosa* (ca. 890 bp, ca. 1100 bp, ca. 1400 bp, Barr et al., 2006). Their comparison with the new ITS1 sequences from this study, was based on semi-quantitative criteria as small differences in indel patterns which may have remained undetected by the previous RFLP analyses, were not taken into account. Patterns of intra- and inter-individual variability of ITS1 were visualized by means of non-metric multidimensional scaling (nMDS) of *p*-distances among sequences.

3. Results

Tajima's test (Tajima, 1989) showed significant *D* values for ITS1 ($D = -2.34$, $p < 0.01$), COI ($D = -1.69$, $0.01 > p > 0.05$) and 16S ($D = -1.66$, $0.01 > p > 0.05$). *D* values were not significant for ND6 ($D = -1.28$) and *per* ($D = -0.93$). Intra-specific sequence divergence of *C. fasciventris* ranged from 0.36% (SE = 0.14%) for 16S to 2.21% (SE = 0.37%) for ND6. *C. anonae* showed levels of intra-specific divergence ranging from 0.10% (SE = 0.06%) for *per* to 0.76% (SE = 0.16%) for ND6. Intra-specific sequence divergence of *C. rosa* ranged from 0.06% (SE = 0.04%) for *per* to 1.25% (SE = 0.28%) for ND6. These values were roughly comparable to inter-specific sequence divergence between morphospecies which varied from 0.03% (SE = 0.02%) to 0.39% (SE = 0.12%) between *C. fasciventris* and *C. anonae* (for 16S and ND6, respectively), from 0.01% (SE = 0.01%) to 1.00% (SE = 0.47%) between *C. fasciventris* and *C. rosa* (for 16S and *per*, respectively) and from 0.03% (SE = 0.01%) to 0.80% (SE = 0.45%) between *C. anonae* and *C. rosa* (for 16S and *per*, respectively).

The number of haplotypes observed for the different markers (Table 3) varied from 13 (16S) to 37 (COI, ITS1). All the five gene fragments yielded haplotypes that were shared among some or all of the three morphospecies. The 16S gene fragment showed haplotypes shared among specimens of (1) *C. fasciventris* (EU276759, EU276760, EU276762, EU276763, EU276766, EU276768, EU276773), *C. anonae* (EU276785, EU276787, EU276789, EU276791, EU276792, EU276795) and *C. rosa* (EU276797, EU276798, EU276799, EU276800, EU276801, EU276804, EU276806, EU276807, EU276808, EU276809, EU276810), (2) *C. fasciventris* (EU276765, EU276767, EU276769, EU276770, EU276771, EU276779) and *C. rosa* (EU276802, EU276803, EU276805, EU276812, EU276813, EU276814), (3) *C. fasciventris* (EU276774, EU276775) and *C. anonae* (EU276794). The COI gene fragment yielded haplotypes shared among specimens of (1) *C. fasciventris* (EU276648, EU276649, EU276652), *C. anonae* (EU276674) and *C. rosa* (EU276684, EU276685, EU276686, EU276690, EU276695, EU276696) and (2). *C. fasciventris* (EU276656) and *C. anonae* (EU276679). The ND6 gene fragment showed haplotypes shared among specimens of (1) *C. fasciventris* (EU276705, EU276706, EU276709) and *C. rosa* (EU276740, EU276741, EU276742, EU276743, EU276747, EU276748, EU276752, EU276753) and (2) *C. fasciventris* (EU276711) and *C. anonae* (EU276730). The *per* gene fragment showed two haplotypes shared among specimens of *C. fasciventris* (EU276829, EU276830, EU276831, EU276832 and EU276833, EU276834, EU276836, EU276837) and *C. anonae* (EU276850, EU276851, EU276852, EU276853 and EU276839, EU276840, EU276841, EU276842, EU276844, EU276845, EU276847, EU276848, EU276849). The ITS1 gene fragment showed two haplotypes shared among specimens of *C. fasciventris* (EU276880 and EU276887, EU276888, EU276891, EU276892) and *C. rosa* (EU276936 and EU276926, EU276929, EU276931, EU276932, EU276937, EU276938, EU276939, EU276941–EU276945, EU276946, EU276947, EU276949, EU276950, EU276955, EU276958, EU276960, EU276962, EU276964–EU276967).

ND6, COI, *per* and ITS1 provided relatively strong genetic signals with a proportion of fully resolved quartets of 73.5%, 71.5%, 77.4% and 94.6%, respectively, while 16S showed only 41.1% fully resolved quartets. None of the five gene fragments recovered monophyletic clades corresponding to the three morphospecies (Figs. 2–5). Accordingly, the SH tests (Table 4) showed significant differences between trees constrained for the monophyly of the morphospecies and topologies of ML trees from COI, ND6, 16S and *per*. ITS1 showed ML tree topologies incompatible with the monophyly of either *C. anonae* and *C. rosa* but congruent with the monophyly of *C. fasciventris*. However, this result should be carefully interpreted as the null hypothesis about the monophyly of *C. fasciventris* was accepted with a significance value only just above the confidence limit ($p = 0.051$).

The COI (Fig. 2) and ND6 (Fig. 3) ML trees grouped the West African specimens of *C. fasciventris* in two bootstrap supported clades, one comprising the specimens from Benin (BS = 99%), the other those from Mali/Ivory Coast (BS = 100% and 99%). *Ceratitis*

Table 3
Alignment information and summary of genetic diversity for specimens of *Ceratitis fasciventris*, *C. anonae* and *C. rosa*

	n	Polym sites	Pars. info sites	No. of haplotypes	π (SD)	*h* (SD)
16S						
C. fasciventris	22	9	3	7	0.003 (0.001)	0.814 (0.049)
C. anonae	16	9	7	7	0.005 (0.001)	0.800 (0.090)
C. rosa	18	2	1	3	0.001 (0.000)	0.542 (0.086)
Total	56	13	9	13	0.003 (0.001)	0.755 (0.049)
COI						
C. fasciventris	22	40	34	18	0.018 (0.001)	0.978 (0.021)
C. anonae	16	22	8	14	0.007 (0.001)	0.983 (0.028)
C. rosa	18	19	9	8	0.007 (0.001)	0.824 (0.072)
Total	56	52	38	37	0.014 (0.001)	0.964 (0.016)
ND6						
C. fasciventris	22	44	30	17	0.021 (0.002)	0.974 (0.022)
C. anonae	16	15	9	12	0.006 (0.001)	0.967 (0.031)
C. rosa	18	27	13	9	0.012 (0.012)	0.012 (0.012)
Total	56	61	41	35	0.016 (0.002)	0.957 (0.019)
per						
C. fasciventris	22	16	9	11	0.008 (0.001)	0.883 (0.053)
C. anonae	16	7	2	6	0.003 (0.001)	0.717 (0.099)
C. rosa	18	10	4	7	0.003 (0.127)	0.634 (0.127)
Total	56	20	8	16	0.008 (0.001)	0.849 (0.027)
ITS1						
C. fasciventris	22	32	14	13	0.014 (0.002)	0.001 (0.039)
C. anonae	30	25	6	13	0.005 (0.097)	0.685 (0.097)
C. rosa	43	52	8	27	0.005 (0.001)	0.940 (0.025)
Total	95	63	23	37	0.008 (0.001)	0.865 (0.026)

n, number of specimens; *h*, haplotype diversity; π, nucleotide diversity.

fasciventris from Kenya and Uganda, on the other hand, was mixed with the two other morphospecies in a large, unresolved clade. The *per* tree (Fig. 4) recovered a large clade comprising all the *C. anonae* and *C. fasciventris* specimens (BS = 90%). Within this clade, subclades with specimens of *C. fasciventris* from Kenya/Uganda (BS = 78%) and from Mali (BS = 97%) were recovered. All the *C. rosa* specimens were placed in a weakly supported clade (BS = 59%). The ITS1 ML tree (Fig. 5) showed a well-supported clade for specimens of *C. fasciventris* from Kenya/Uganda (BS = 86%). A second large clade included specimens from all the three morphospecies (BS = 95%). The 16S ML tree showed generally low bootstrap values and large unresolved polytomies (data not shown). According to the SH tests, tree topologies produced by COI, ND6, 16S and *per* were significantly different (Table 5).

Cloning of the ITS1 gene fragment in 21 FAR specimens produced 95 sequences of three different length classes (hereafter group A, B and C). Group A included 61 sequences of ca. 1100 bp (EU276880, EU276883–EU276894, EU27689–EU276902, EU276925–EU276967) belonging to specimens of *C. fasciventris* from Tanzania ($n = 1$), Mali ($n = 1$), Ivory Coast ($n = 2$) and Benin ($n = 4$), of *C. anonae* from Ghana ($n = 2$) and of *C. rosa* from Tanzania ($n = 4$), Reunion ($n = 4$) and South Africa ($n = 6$). Group B included 32 sequences of ca. 890 bp (EU276873–EU276879, EU276895–EU276897, EU276903–EU276924) belonging to specimens of *C. fasciventris* from Kenya ($n = 4$) and Uganda ($n = 2$), of *C. anonae* from Kenya ($n = 6$), Uganda ($n = 4$), Cameroon ($n = 3$), Benin ($n = 3$), Ghana ($n = 1$) and Ivory Coast ($n = 3$). Group C included two sequences of ca. 770 bp (EU276881, EU276882) belonging to a single specimen of *C. fasciventris* from Benin. Interestingly this specimen of *C. fasciventris* produced also an ITS1 sequence belonging to group A (EU276880). The alignment of 95 FAR ITS1 sequences resulted in 37 haplotypes (not considering sites with gaps/missing data) with an average sequence divergence of 0.8% (SE = 0.2%). Levels of intra-individual variability were relatively high with an average sequence divergence within specimens of 0.5% (SE = 0.02%). Up to seven different haplotypes were found within a single specimen differing either in base composition and/or indels. The occurrence of more than two ITS1 alleles proves that intra-individual variability in the FAR complex is not only a consequence of heterozygosis but is due to the presence of multiple ITS1 repeat copies in the genome. The lack of evident clustering in the nMDS (Fig. 6), shows that values of intra-individual, intra- and inter-specific *p*-distances for ITS1 are roughly comparable.

4. Discussion

The present analysis using specimens from all over the distributional area of the FAR complex and surveying five different gene fragments, did not recover *C. fasciventris*, *C. anonae* and *C. rosa*, as three well-defined clades. All the markers provided phylogenetic signals incongruent with the classification based on morphological characters (De Meyer, 2001b, 2005; De Meyer and Freidberg, 2006). Levels of intra-specific genetic divergence were generally comparable or even higher than inter-specific differences. This was particularly evident (a) for ND6 and COI, where the genetic differences among samples of *C. fasciventris* were markedly higher than inter-specific differences among the three morphospecies and (b) for ITS1, where intra-specific and intra-individual sequence differences were comparable to, or even larger than, the inter-specific sequence divergences.

Likelihood mapping and ML trees suggested that the strength of phylogenetic signals provided by COI, ND6, *per* and ITS1 is adequate for phylogenetic inferences (see Strimmer and von Haeseler, 1997). Conversely, the genetic signal provided by 16S was low and heavily affected by phylogenetic noise. Hence, sequence analyses of 16S seem inadequate to resolve the shallow phylogenies of the FAR complex and should not be considered either for the molecular characterization of these species or as a possible alternative to the standard DNA barcoding region (Hebert, 2003).

Tajima's test (Tajima, 1989) produced significant *D* values for three out of five gene fragments analysed. Among them, ITS1 showed highly significant *D* values, even if selective constraints acting on this intergenic spacer are supposed to be minimal (Parkin and Butlin, 2004). Tajima's *D* statistic is widely used in testing the

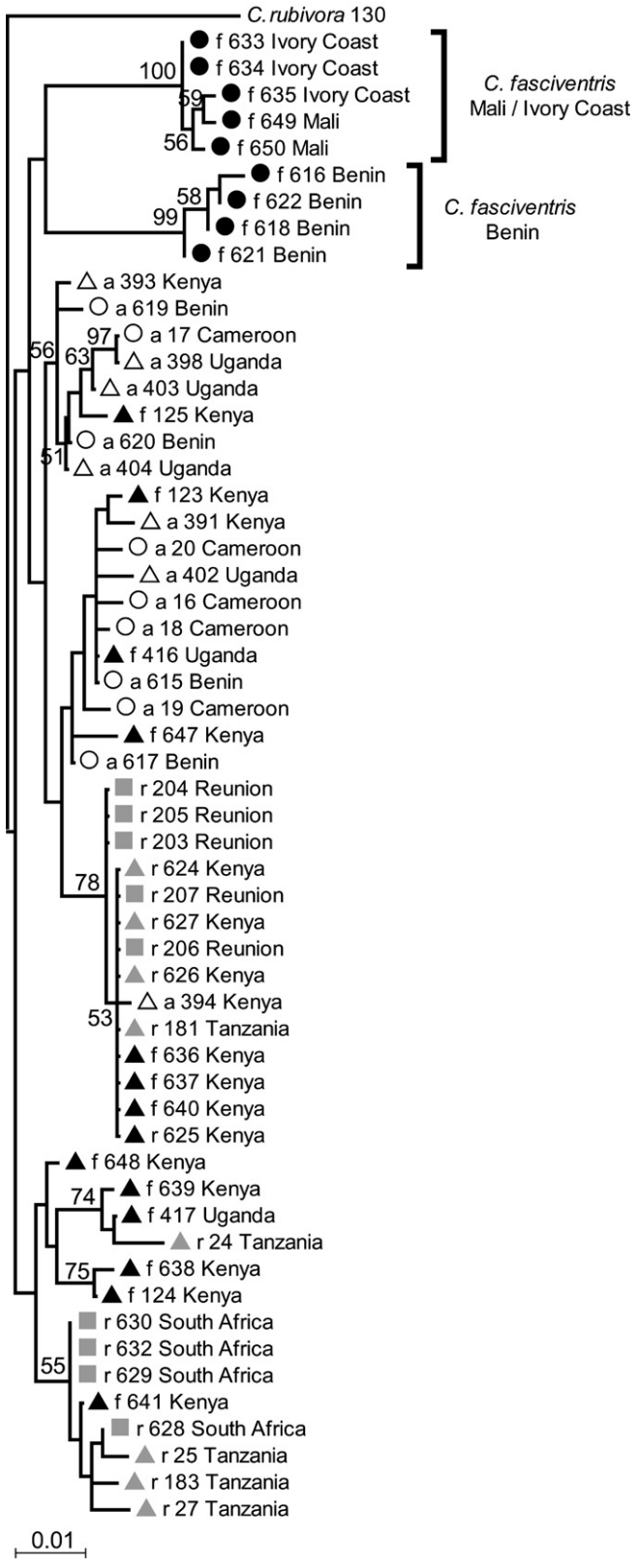

Fig. 2. Maximum likelihood tree of the COI gene. Black, white and grey labels indicate *C. fasciventris*, *C. anonae* and *C. rosa*, respectively. Triangles, circles and squares were used for Eastern, Western and Southern locations, respectively (see Table 1). Only bootstrap values above 50% are shown.

neutral equilibrium model. The null hypothesis is a composite hypothesis that includes assumptions regarding the demographics of the populations, such as constant population size and lack of population structuring. As previously reviewed by Nielsen (2001) one of the limitations of this and other neutrality tests (e.g. Fu and Li, 1993) is that both demographic process and selection can have similar effects on gene genealogies. Hence, the significant *D* values observed in this study may be indicative of genetic substructuring (as observed among populations of *C. fasciventris*), rather than of selective processes.

COI and ND6 showed the occurrence of two strongly supported clades of *C. fasciventris* from Benin and Mali/Ivory Coast. This pattern was not recovered by *per* and ITS1 which, in turn showed a well-supported clade corresponding to *C. fasciventris* from Kenya/Uganda. Due to the incongruence among gene genealogies obtained with the different markers (and in particular among the mitochondrial and the nuclear gene genealogies) and to the difference in the strength of phylogenetic signals we did not analyse the concatenated dataset (see Edwards et al., 2007). Conversely, comparisons of nuclear and mitochondrial DNA genealogies can provide useful indications on the evolutionary processes affecting shallow phylogenies (Shaw, 2002). Recombination and backcrossing could affect different parts of the genome resulting in incongruent phylogenies of DNA fragments isolated from a single individual (Sota and Vogler, 2001). Moreover, mitochondrial and nuclear loci may differ in their susceptibility to evolutionary processes (Moore, 1995; Funk and Omland, 2003). Accordingly, the incongruence of gene genealogies between the nuclear loci *per* and ITS1 and the mitochondrial loci COI and ND6 suggests that patterns of genetic divergence within the complex may involve processes such as incomplete lineage sorting and/or introgressive hybridisation. The highly polyphagous species of the FAR complex show largely overlapping distributional areas and host spectra (Copeland et al., 2006). However, the elaborate species-specific mating behaviour of fruit flies (Quilici et al., 2002; Pike et al., 2003) is supposed to prevent inter-specific hybridisation under field conditions. Interspecific hybrids of the three FAR morphospecies have hitherto not been reported from natural populations. However, this may be not surprising if hybridisation is rare (Feder et al., 1999; Pike and Meats, 2002). Moreover, recent laboratory experiments succeeded in producing hybrids from different combinations of *C. fasciventris*, *C. anonae* and *C. rosa* (Erbout et al., 2008) suggesting that the reproductive barriers among the three species may be less effective than previously hypothesised.

The use of ITS1 to identify species of the FAR complex still appears controversial. This study suggests that under particular circumstances, ITS1 length differences may separate *C. anonae* and *C. fasciventris* on one hand, from C. *rosa* on the other (the latter not producing ca. 890 bp fragments). Furthermore, Douglas and Haymer (2001) reported amplicon size differences between *C. rosa* from South Africa and Kenya. Yet, they did not distinguish between *C. rosa* and *C. fasciventris*, as this latter taxon acquired the species status only afterward (De Meyer, 2001b). This is probably why they interpreted the ITS1 length polymorphism as intra-specific variation. Barr et al. (2006), analysing ITS1 in a larger number of samples and specimens suggested that the ITS1 length difference could separate *C. rosa* from *C. fasciventris* and *C. anonae* (*C. rosa* producing ca. 1100 bp and ca. 1400 bp fragments, *C. fasciventris* and *C. anonae* only ca. 890 bp fragments). The present study, however, showed that indel patterns of AF189690 (ca. 1100 bp) were markedly similar to those of ITS1 sequences from group A (including specimens of *C. fasciventris*, *C. anonae* and *C. rosa*), while indels of DQ645954 and AF189689 (ca. 890 bp) were comparable to those of sequences from group B (including specimens of both *C. fasciventris* and *C. anonae*). None of the ITS1 sequences analysed showed the long insertion of DQ645953 (ca. 1400 bp). It remains to be clarified whether this sequence represents an intra-specific variant of populations of *C. rosa* from Malawi and Kenya or involves still another cryptic taxon. Anyway, these results suggest that tax-

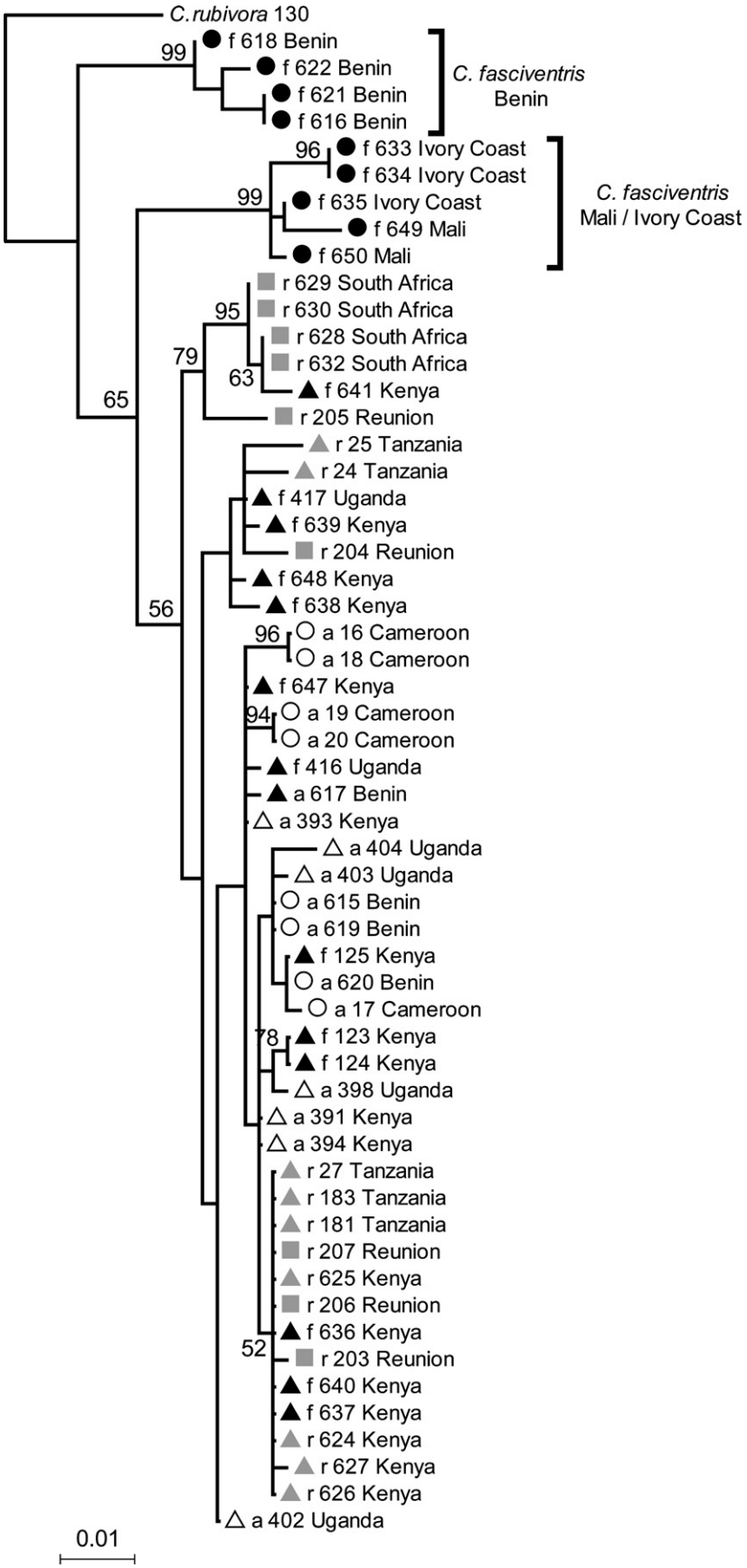

Fig. 3. Maximum likelihood tree of the ND6 gene. Black, white and grey labels indicate *C. fasciventris*, *C. anonae* and *C. rosa*, respectively. Triangles, circles and squares were used for Eastern, Western and Southern locations, respectively (see Table 1). Only bootstrap values above 50% are shown.

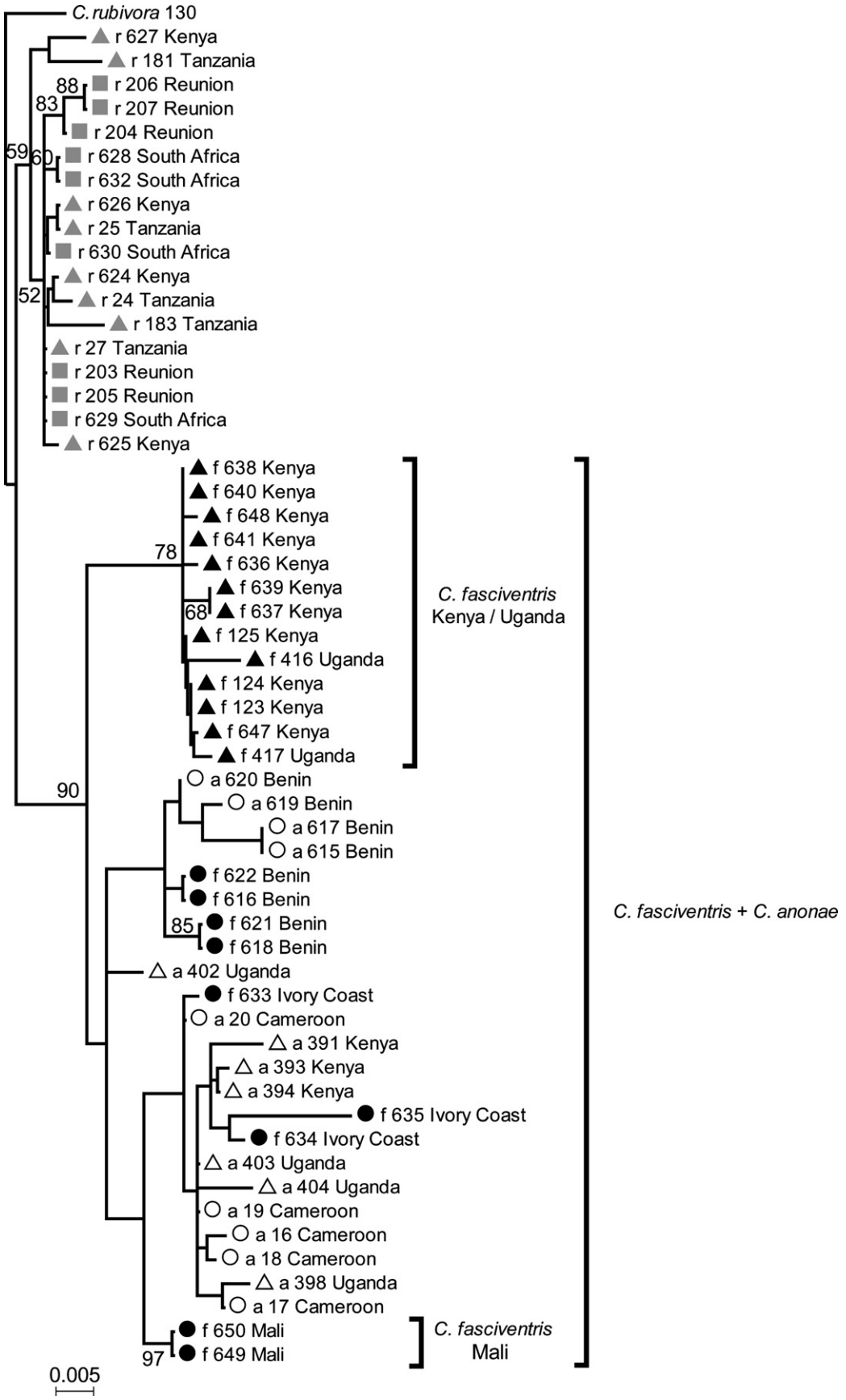

Fig. 4. Maximum likelihood tree of the *per* gene. Black, white and grey labels indicate *C. fasciventris*, *C. anonae* and *C. rosa*, respectively. Triangles, circles and squares were used for Eastern, Western and Southern locations, respectively (see Table 1). Only bootstrap values above 50% are shown.

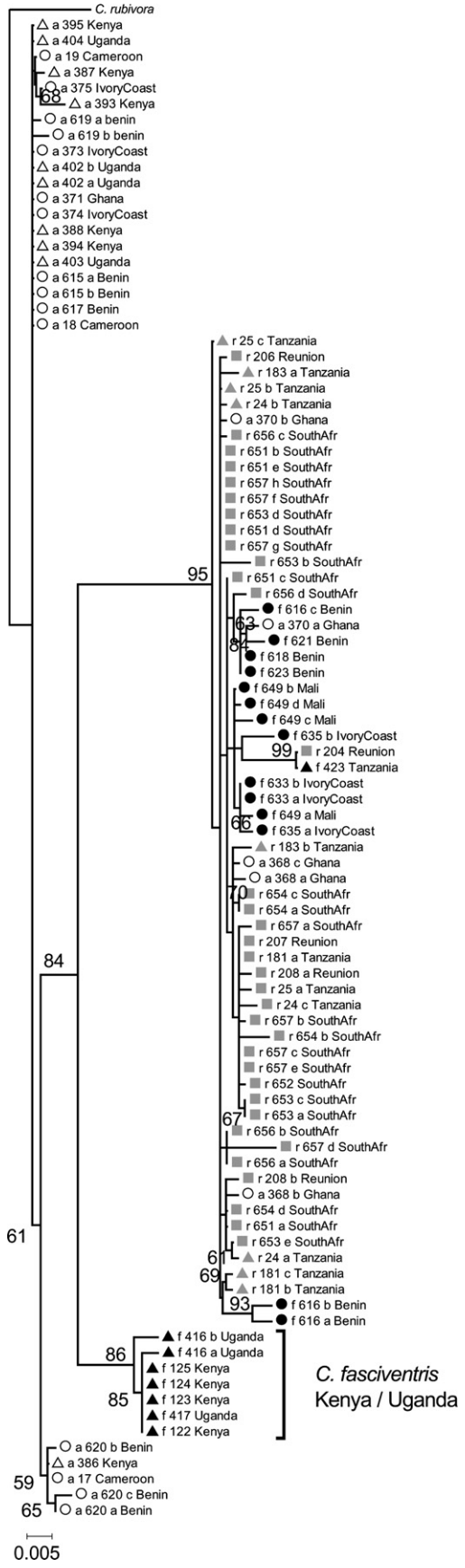

Fig. 5. Maximum likelihood tree of the ITS1 gene. Black, white and grey labels indicate *C. fasciventris*, *C. anonae* and *C. rosa*, respectively. Triangles, circles and squares were used for Eastern, Western and Southern locations, respectively (see Table 1). Only bootstrap values above 50% are shown.

Table 4
Results of Shimodaira–Hasegawa tests comparing likelihood scores of unconstrained majority rule consensus trees (bootstrap support ⩾ 70%) and trees constrained for the monophyly of *C. fasciventris*, *C. anonae* or *C. rosa* for the five markers investigated (COI, ND6, 16S, ITS1, *per*)

Dataset	Constrain	No. of compatible trees	$-\ln L$	Diff. $-\ln L$	*p* value
16S	None		1169.3		
	C. fasciventris	82	1029.2	140.0	**0.000**
	C. anonae	411	1014.3	155.0	**0.000**
	C. rosa	162	1008.4	160.8	**0.000**
COI	None		2031.4		
	C. fasciventris	1	1872.0	159.5	**0.000**
	C. anonae	4	1836.8	194.6	**0.000**
	C. rosa	1	1844.7	186.8	**0.000**
ND6	None		1610.6		
	C. fasciventris	1	1456.7	153.9	**0.006**
	C. anonae	1	1408.8	201.8	**0.002**
	C. rosa	1	1440.4	170.2	**0.005**
per	None		1763.5		
	C. fasciventris	147	1624.5	139.0	**0.002**
	C. anonae	3	1617.3	146.2	**0.001**
	C. rosa	26	1602.6	160.9	**0.001**
ITS1	None		2794.4		
	C. fasciventris	2	2397.7	396.7	**0.051**
	C. anonae		2381.7	412.7	**0.030**
	C. rosa	2	2394.4	399.9	**0.040**

Significant *p*-values ($p < 0.05$) after the False Discovery Rate correction (Benjamini and Hochberg, 1995) are in bold.

Table 5
Results of Shimodaira–Hasegawa tests comparing ML tree topologies of different markers

	Constrain	16S	COI	ND6	*per*
Dataset					
16S		–	**0.007**	**0.042**	**0.002**
COI		120.1 (1)	–	**0.000**	**0.001**
ND6		119.0 (2)	160.4 (1)	–	**0.024**
per		127.4 (3)	139.3 (3)	81.9 (2)	–

Below diagonal: difference in likelihood scores between unconstrained and constrained trees (in parentheses number of compatible trees). Above diagonal: *p*-values. Significant *p*-values ($p < 0.05$) after the False Discovery Rate correction (Benjamini and Hochberg, 1995) are in bold. Majority rule consensus trees (bootstrap support ⩾ 70%) were used for the comparisons.

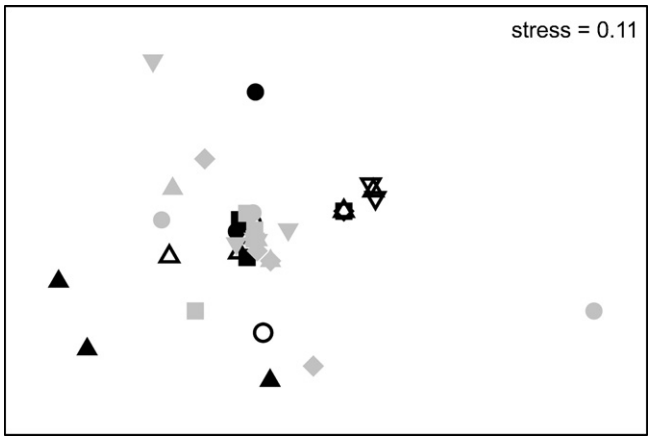

Fig. 6. Non-metric multidimensional scaling (nMDS) plot showing the relative amount of intra-individual, intra- and inter-specific genetic divergence among ITS1 sequences. nMDS is based on *p*-distances between sequences of *C. fasciventris* (black labels), *C. anonae* (white labels) and *C. rosa* (grey labels). Clones obtained from the same specimen are represented with the same shape and colour.

on sampling may heavily affect estimates of intraspecific ITS1 variability and that ITS1 length polymorphism does not represent a

foolproof diagnostic tool to discriminate *C. fasciventris*, *C. anonae* and *C. rosa* across their entire distributional range.

An intriguing perspective is the possibility of recognizing site-specific differences (e.g. SNPs) allowing to discriminate one or more species of the FAR complex. Such diagnostic characters, however, did not appear evident after a first survey of the dataset, and could have remained undetected by using the gene-tree approach adopted in this study. Further analyses are needed to evaluate this possibility.

In conclusion, the molecular data obtained so far do not allow to identify *C. fasciventris*, *C. anonae* and *C. rosa* as three distinct phylogenetic species (as defined in Mishler and Theriot, 2000). The phylogenetic relationships and taxonomic interpretation of *C. fasciventris*, *C. anonae* and *C. rosa* appear more complex than previously suggested, with incongruence either among molecular markers and between molecular markers and taxonomical status. The non-monophyly of the three morphospecies, the occurrence of shared haplotypes and the possibility of inter-specific hybridisation under laboratory conditions are consistent with two not mutually exclusive models: (a) the taxa might be (partly) reproductively isolated but diverged only recently such that insufficient time has passed for neutral DNA to differentiate by genetic drift and (b) the taxa may largely act as species but occasionally cross-bred such that gene flow results in genetic homogenization of part of the genome. The scenario emerging from the analysis of East and West African locations suggests that *C. fasciventris* might be itself a complex of cryptic taxa. Ongoing analyses seem to confirm morphological differences in males of *C. fasciventris* from West African locations as previously mentioned in De Meyer (2001b). However, the occurrence of genetically isolated populations of *C. fasciventris* and of effective genetic barriers among sympatric populations of *C. fasciventris* and either *C. anonae* or *C. rosa* needs to be further investigated through the analysis of an adequate number of samples and specimens.

Acknowledgments

We wish to thank Bruno Nevado for his valuable support during the different stages of this study. We are grateful to Erik Verheyen and the discussion group of the Molecular Lab of RBINS for their stimulating comments on earlier drafts of the paper. We thank the various colleagues who provided specimens for this study. This study benefited from practical support by the "Joint Experimental Molecular Unit (JEMU)" of RBINS and RMCA.

References

Baliraine, F.N., Bonizzoni, M., Guglielmino, C.R., Osir, E.O., Lux, S.A., Mulaa, F.J., Gomulski, L.M., Zheng, L., Quilici, S., Gasperi, G., Malacrida, A.R., 2004. Population genetics of the potentially invasive African fruit fly species, *Ceratitis rosa* and *Ceratitis fasciventris* (Diptera: Tephritidae). Mol. Ecol. 13, 683–695.

Barr, N.B., Copeland, R.S., De Meyer, M., Masiga, D., Kibogo, H.G., Billah, M.K., Osir, E., Wharton, R.A., McPheron, B.A., 2006. Molecular diagnostics of economically important fruit fly species (Diptera: Tephritidae) in Africa using PCR and RFLP analyses. Bull. Entomol. Res. 96, 505–521.

Barr, N.B., McPheron, B.A., 2006. Molecular phylogenetics of the genus *Ceratitis* (Diptera: Tephritidae). Mol. Phylogenet. Evol. 38, 216–230.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Stat. Soc. B 57, 289–300.

Copeland, R.S., Wharton, R.A., 2006. Year-round production of pest *Ceratitis* species (Diptera: Tephritidae) in fruit of the invasive species *Solanum mauritianum* in Kenya. Ann. Entomol. Soc. Am. 99, 3530–3535.

Copeland, R.S., Wharton, R.A., Luke, Q., De Meyer, M., Lux, S., Zenz, N., Machera, P., Okumu, M., 2006. Geographic distribution, host fruit, and parasitoids of African fruit fly pests *Ceratitis anonae*, *Ceratitis cosyra*, *Ceratitis fasciventris*, and *Ceratitis rosa* (Diptera: Tephritidae) in Kenya. Ann. Entomol. Soc. Am. 99, 261–278.

De Meyer, M., 2001a. Distribution patterns and host–plant relationships within the genus *Ceratitis* MacLeay (Diptera: Tephritidae) in Africa. Cimbebasia 17, 219–228.

De Meyer, M., 2001b. On the identity of the Natal fruit fly *Ceratitis rosa* Karsch (Diptera, Tephritidae). Bull. Inst. Roy. Sci. Nat. Belg. Entomol. 71, 55–62.

De Meyer, M., 2005. Phylogenetic relationships within the fruit fly genus *Ceratitis* MacLeay (Diptera: Tephritidae), derived from morphological and host plant evidence. Insect. Syst. Evol. 36, 459–480.

De Meyer, M., Copeland, R.S., Lux, S., Mansell, M., Wharton, R., White, I.M., Zenz, N., 2002. Annotated check list of host plants for Afrotropical fruit flies (Diptera: Tephritidae) of the genus *Ceratitis*. Docum. Zool. Mus. r. Afr. cent. 27, 1–92.

De Meyer, M., Freidberg, A., 2006. Revision of the subgenus *Ceratitis* (*Pterandrus*) Bezzi (Diptera: Tephritidae). Isr. J. Entomol. 36, 197–315.

De Meyer, M., Robertson, M.P., Peterson, A.T., Mansell, M.W., 2008. Ecological niches and potential geographical distributions of Mediterranean fruit fly (*Ceratitis capitata*) and Natal fruit fly (*Ceratitis rosa*). J. Biogeogr. 35, 270–281.

Douglas, L., Haymer, D.S., 2001. Ribosomal ITS1 polymorphisms in *Ceratitis capitata* and *Ceratitis rosa* (Diptera: Tephritidae). Ann. Entomol. Soc. Am. 94, 726–731.

Duyck, P.-F., David, P., Quilici, S., 2004. A review of relationships between interspecific competition and invasions in fruit flies (Diptera: Tephritidae). Ecol. Entomol. 29, 511–520.

Edwards, S.V., Liu, L., Pearl, D.K., 2007. High-resolution species trees without concatenation. Proc. Natl. Acad. Sci. USA 104, 5936–5941.

Erbout, N., De Meyer, M., Lens, L., 2008. Hybridisation between two polyphagous fruit fly species (Diptera: Tephritidae) causes sex-biased reduction in developmental stability. Biol. J. Linn. Soc. 93, 579–588.

Feder, J.L., Williams, S.M., Berlocher, S.H., McPheron, B.A., Bush, G.L., 1999. The population genetics of the apple maggot fly, *Rhagoletis pomonella* and the snowberry maggot, *Rzephyria*: implications for models of sympatric speciation. Entomol. Exp. Appl. 90, 9–24.

Fu, Y.X., Li, W.H., 1993. Statistical tests of neutrality of mutations. Genetics 133, 693–709.

Funk, D.J., Omland, K.E., 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annu. Rev. Ecol. Syst. 34, 397–423.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. 41, 95–98.

Hancock, D.L., 1989. Pest status; Southern Africa (subfamily Ceratitidinae). In: Robinson, A., Hooper, G. (Eds.), Fruit Flies, Their Biology, Natural Enemies and Control. Elsevier, New York, pp. 54–58.

Hancock, D.L., White, I.M., 1997. The identity of *Tririthrum nigrum* (Graham) and some new combinations in *Ceratitis* MacLeay (Diptera: Tephritidae). Entomologist 116, 192–197.

Hasegawa, M., Kishino, H., Yano, T., 1985. Dating the human–ape split by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22, 160–174.

Hebert, P.D.N., 2003. Biological identifications through DNA barcodes. Proc. R. Soc. Lond. B 270, 313–321.

Hillis, D.M., Bull, J.J., 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42, 182–192.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinform. 5, 150–163.

Malacrida, A., Gomulski, L., Bonizzoni, M., Bertin, S., Gasperi, G., Guglielmino, C., 2007. Globalization and fruit fly invasion and expansion: the medfly paradigm. Genetica 131, 1–9.

Mishler, B.D., Theriot, E., 2000. The phylogenetic species concept *sensu* Mishler and Theriot: monophyly, apomorphy, and phylogenetic species concepts. In: Wheeler, Q.D., Meier, E. (Eds.), Species Concepts and Phylogenetic Theory: A Debate. Columbia University Press, New York, pp. 44–54.

Moore, W.S., 1995. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. Evolution 49, 718–726.

Nielsen, R., 2001. Statistical tests of selective neutrality in the age of genomics. Heredity 86, 641–647.

Nylander, J.A.A., 2004. MrModeltest v2. Program Distributed by the Author. Evolutionary Biology Centre, Uppsala University.

Parkin, E.J., Butlin, R.K., 2004. Within- and between-individual sequence variation among ITS1 copies in the meadow grasshopper *Chorthippus parallelus* indicates frequent intrachromosomal gene conversion. Mol. Biol. Evol. 21, 1595–1601.

Pike, N., Meats, A., 2002. Potential for mating between *Bactrocera tryoni* (Froggatt) and *Bactrocera neohumeralis* (Hardy) (Diptera: Tephritidae). Aust. J. Entomol. 41, 70–74.

Pike, N., Wang, W.Y.S., Meats, A., 2003. The likely fate of hybrids of *Bactrocera tryoni* and *Bactrocera neohumeralis*. Heredity 90, 365–370.

Quilici, S., Franck, A., Peppuy, A., Dos Reis Correia, E., Mouniama, C., Blard, F., 2002. Comparative studies of courtship behaviour of *Ceratitis* spp. (Diptera: Tephritidae) in Reunion island. Florida Entomol. 85, 138–142.

Rozas, J., Sanchez-DelBarrio, J.C., Messeguer, X., Rozas, R., 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19, 2496–2497.

Schmidt, H.A., Strimmer, K., Vingron, M., von Haeseler, A., 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18, 502–504.

Schneider, S., Roessli, D., Excoffier, L., 2000. Arlequin, Version 2.000: A Software for Population Genetics Data Analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Shaw, K.L., 2002. Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. Proc. Natl. Acad. Sci. USA 99, 16122–16127.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114–1116.
Sota, T., Vogler, A.P., 2001. Incongruence of mitochondrial and nuclear gene trees in the carabid beetles *Ohomopterus*. Syst. Biol. 50, 39–59.
Strimmer, K., von Haeseler, A., 1997. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. Proc. Natl. Acad. Sci. USA 94, 6815–6819.
Swofford, D.L., 2002. Phylogenetic Analysis Using Parsimony (and Other Methods). Version 4. Sinauer Associates, Sunderland, MA.
Tajima, F., 1989. Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. Genetics 123, 585–595.
Tavaré, S., 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. In: Miura, R.M. (Ed.), Some Mathematical Questions in Biology—DNA Sequence Analysis. American Mathematical Society, Providence, pp. 57–86.